



APPENDIX-V

Test with Direct Coding of Words

In these tests, patterns of holes were used to record the letters used to spell out English words of translated titles of Russian scientific, technical, and economics articles. The patterns of holes spelled out the words in exactly the same way that embossed patterns of dots spell out words in Braille.

25X1A9a

Experiments of this type were first suggested by [REDACTED] of OSI. The following considerations provided the impetus to these tests. When writing for publication, scientists usually devote considerable care to providing titles that indicate at least the more important aspects of the subject matter of their papers. The number of synonyms used to designate these general aspects of papers in a given field is usually fairly restricted in number. Hence a machine search directed to the words used in titles can provide the basis for rapid screening of such papers.

25X1A9a

A particularly important practical consideration is the fact that there exist in the Agency several extensive files of titles for which no index nor other means for selecting out papers relating to a given subject are available. Exploiting the material in these files has been a problem for which no effective and economically feasible procedure has been developed up to the present. Even if machine searching based on words in the titles would result, on the average, in locating 50% of the papers pertinent to a given subject, the exploitation of these files would be on a much more effective basis than has been possible in the past. Such exploitation would probably provide much information badly needed for intelligence purposes.

Three sets of tests have been initiated.

I. CHEMISTRY

An IBM card file was prepared of 1300 titles of Russian articles abstracted in "Chemical Abstracts" in the field of physical chemistry with one title punched in each card. A typical card was punched as follows:

0491 extract of linseed meal as an inhibitor of iron
and steel corrosion

The number (0491) is the arbitrarily assigned serial number of a card on which the author, title, journal reference, and abstract are recorded. In some cases the titles of the articles exceeded

~~CONFIDENTIAL~~
Security Information

the capacity of the 80 column IBM card; these were shortened by the key punch operator by the omission of less significant words such as "of", "and", "the", and the like.

The first significant searching experiment was conducted as follows:

25X1A9a

██████████ formulated a question relating to the activity of USSR chemists working in the field of protective films on metals. He visually scanned the 1300-odd titles of the file and selected 108 titles which he felt to be pertinent to this problem. The words appearing in these titles were noted and a machine search was then directed to the IBM file. The machine was instructed to select those titles which contained any one of the following terms:

corrosion
anticorrosion
film
protective
coating
electrode
diffusion

Each word in every title was scanned automatically by the machine in about 3 minutes. All titles were selected in which any one of the seven key words appeared. The machine selected 125 titles of which 105 were included in the 108 previously manually selected by Dr.

25X1A9a

██████████ the remaining 20 titles were considered to be nonpertinent.

Further machine searches directed to other questions and to other facets of the above question indicated that the results presented above are somewhat better than might ordinarily expected. However, it may be concluded that rough, though effective screening of a title file may be accomplished with minimum expenditure of time and effort. Additional runs are now being made in order to obtain more quantitative data, to estimate the effectiveness of application of the scanning equipment.

II. MEDICINE

A second IBM card file of some 1800 translated titles of published Russian medical articles was also prepared. In this file, the letters were punched as two-hole patterns ordinarily used with standard equipment (so-called "standard" punching).

In one test with this file, a preliminary search to the word "Dysentery" was carried out. The selected cards were subjected to a second search to any one of the key words "Prevention", "Therapy", "Antibiotics", "Treatment", "Sulfo" (as in "Sulfo thiazine" for example). The two sets of selected cards are being examined by Dr. ██████████ and are being checked against the rejected cards to est-

25X1A9a

~~CONFIDENTIAL~~
Security Information

-2-

~~CONFIDENTIAL~~
Security Information

imate the effectiveness of this form of machine searching.

In a second search, it was desired to select titles of papers relating to the physiology of the central nervous system. The machine was set to select titles containing either "nerv-" (as in "nerve" or "nervous") or "vasc-" (as in "vascular") or "neur-" as in "neural". The titles selected included a number of papers not relating directly to the scope of search as originally proposed. In spite of this, the screening accomplished was deemed highly effective.

III. GENERAL

25X1A9a

25X1A9a

At the suggestion of [REDACTED], ORR, and after consultation with Miss [REDACTED] OCD, a third experimental file capable of being scanned automatically was prepared using titles derived from the "Monthly List of Russian Accessions". The fields of interest represented in this file of about 2000 titles cover most of science and technology as well as economic interests. It is expected that experimental searching will have been initiated by the time this report is being distributed.

CONCLUSION

Although results to date scarcely warrant positive recommendations, it seems likely that this mode of employing automatic scanning equipment may prove valuable in extracting useful information from otherwise dormant sources.

It is important to note that preparation of files of the type described above require no more than clerical work to punch cards. Eventually it should be possible to greatly improve the effectiveness of searching based on titles by converting their component terms into codes based on semantic factoring. A simple translating machine will permit this operation to be performed with the same clerical effort now required to punch words in cards. A clerk trained to recognize Russian letters could work directly from untranslated Russian titles.

~~CONFIDENTIAL~~
Security Information